

CA1

BS

-1990

G743

C.2

Government
Publications

CA 1
BS
- 1990
G 743

Government
Publications

Guide to Literacy Levels on the Survey of Literacy Skills Used in Daily Activities

Prepared by
Stan Jones
Survey Consultant

May 30, 1990

2
1

Contents

1	Introduction	1
2	A functional literacy continuum	1
3	Test items for a functional literacy continuum	3
3.1	Designing reading items	3
3.2	Administering the test	3
4	Measuring functional literacy	4
4.1	Determining how to measure levels	4
4.2	Assigning individuals to levels	5
4.2.1	Checking reliability	5
4.2.2	Grouping items by difficulty	5
4.2.3	Determining an individual's level	6
5	Conclusion	7
6	Notes	9
7	Bibliography	10

Tables

Table 1	Definitions of levels of reading skill	2
Table 2	Percentage distribution of Canadian adults aged 16-69 by level of reading skills.	7



Digitized by the Internet Archive
in 2023 with funding from
University of Toronto

<https://archive.org/details/39192516010352>

Guide to Literacy Levels on the Survey of Literacy Skills Used in Daily Activities

Stan Jones
May 30, 1990

1. Introduction

It is generally agreed that functional literacy skills do not neatly fall into categories, but rather form a continuum:

It seems more appropriate to represent functional literacy as continuously distributed, with various points along the continuum indicating different levels of functioning. (Kirsch & Guthrie, 1981)

In reporting the results of the Survey of Literacy Skills Used in Daily Activities (LSUDA), the national survey of everyday literacy skills in Canada conducted by Statistics Canada for the National Literacy Secretariat, it is necessary to recognize this fact about literacy. At the same time it is important, for program and policy needs, to mark certain points or levels along the continuum as worthy of particular attention. The levels used in the design of the survey and in reports on it are simply points along the functional literacy continuum that we believe will be useful to governments in identifying types of programs needed to deal with the literacy problem and to literacy providers in identifying clients, possibly new kinds of clients, for their services. We also think that these points reflect significant differences in literacy abilities. This paper discusses the rationale and the process for identifying these points.

The next section of the paper begins with some further discussion of functional literacy as a continuum. It then lays out how the levels were defined. The third section discusses how a test of functional literacy was designed and administered. The fourth part of the paper discusses how the test was scored to measure the literacy levels. The paper concludes with a brief presentation of some results from the test.

2. A functional literacy continuum

If we are to identify points that are of interest in this way, it is important to rely on an adequate theory of functional literacy, particularly as it pertains to functional reading. Although there is no fully comprehensive theory as yet, enough is known through the work of Mikulecky on the task context of functional literacy (Mikulecky, 1985) and through the work of Kirsch and Guthrie on the cognitive differences between school and functional reading (Guthrie, 1988; Guthrie & Kirsch, 1987) to permit us to proceed.

These studies suggest that functional literacy is dependant both on the ability to decode relatively small and not necessarily connected chunks of text and on the knowledge of how to apply the information gained to solve a problem. Thus, any

continuum of functional literacy must take into account both decoding and decision criteria. The framework used in the LSUDA survey have been thoroughly discussed in the planning document (Kelly, Satin & Murray, 1987) and in reports on the project (Jones, 1989; Jones, Satin, Kelly, & Montigny, 1990). The planning document definition of the levels has remained the guide throughout the project, but as the project team discussed these levels among themselves and with others a more precise way of describing them has been developed.

As we will be concerned principally with the measurement of document reading in this report, only that continuum is reprinted here as Table 1.

What is crucial to note is that the points along the continuum were developed prior to the development of any test items and served to guide that development. For example, items that were to distinguish between level 1 and level 2 abilities were designed to require only the ability to recognize and point out key words or short phrases in the text. It was not assumed that all these items would be equally easy or difficult; other factors that we could not measure, such as prior familiarity with the information in the text, could influence any individual's response. However, if we provided enough items at this level and used an analysis procedure that was sensitive to the overall pattern of responses we would minimize the effect of individual variation¹.

Table 1 Definitions of levels of reading skill

LEVEL	DESCRIPTION
1	Canadians at this level have difficulty dealing with printed materials. They most likely identify themselves as people who cannot read.
2	Canadians at this level can use printed materials for limited purposes only, such as finding a familiar word in a simple text. They would likely recognize themselves as having difficulties with common reading materials.
3	Canadians at this level can use reading materials in a variety of situations, provided the material is simple, clearly layed out and the tasks involved are not too complicated. While these people generally do not see themselves as having significant reading difficulties, they tend to avoid situations requiring reading.
4	Canadians at this level meet most everyday reading demands. This is a diverse group which exhibits a wide range of reading skills.

1	Canadians at this level have difficulty dealing with printed materials. They most likely identify themselves as people who cannot read.
2	Canadians at this level can use printed materials for limited purposes only, such as finding a familiar word in a simple text. They would likely recognize themselves as having difficulties with common reading materials.
3	Canadians at this level can use reading materials in a variety of situations, provided the material is simple, clearly layed out and the tasks involved are not too complicated. While these people generally do not see themselves as having significant reading difficulties, they tend to avoid situations requiring reading.
4	Canadians at this level meet most everyday reading demands. This is a diverse group which exhibits a wide range of reading skills.

The levels, then, are not points derived after the fact from the test data, but were actually designed into the test. Thus the LSUDA survey results do not provide data to discover what the points/levels were, but rather data to confirm or disconfirm the model of functional literacy, reflected in the levels, that originally generated the specific test items.

3. Test items for a functional literacy continuum

3.1 Designing reading items

The LSUDA survey was to be a direct assessment; that is, the measure of an individual's level of ability was to be a test of that individual's ability to carry out tasks at a particular level of difficulty. Further, the items on the test were to require the individual to read a real text, not one made up for the test, and use the information in a realistic way. In other words, the test items were to simulate real life reading tasks.

In designing items of this kind for the LSUDA test, whether a new item or one adapted from existing items used in other functional literacy surveys, our intent was to create a pool of items that would serve as tests of the various levels. Thus, the pool of level 2 items would be those that only require the individual to locate key words or short phrases in a text. The best example of such an item at level 2 is one which required individuals to look over a grocery shopping list and identify all the items on the list that were in an ad from a supermarket. Other level 2 items were either a bit simpler - one required individuals to identify a particular sign - or a bit more complex - finding where to vote from an enumeration form, but all focused on finding key words.

A typical level 3 item required individuals to find out when they had to return a form to their child's school. Rather than simply find a word, they had to understand a sentence and decide what action it required. A more complex item at this level had individuals find out whether a particular kind of sandpaper could be used for a particular job.

As in all tests of this kind, the items were designed to fit the specifications - the reading levels - so that the abilities of individuals could be measured. Parallel items were developed in English and French.

3.2 Administering the test

The LSUDA survey was administered to a sample of adult Canadian aged 16-69 who were selected from households that had been surveyed as part of Statistics Canada's Labour Force Survey (LFS) in the previous six months. The LFS is a monthly household survey of labour force activity using a large representative sample drawn from households across Canada. Residents of the Yukon and the Northwest Territories, members of the Armed Forces, persons living

on Indian reserves and inmates of institutions were not included in the sample. These exclusions account for approximately 3% of the Canadian population. Although LFS samples households, LSUDA surveyed individuals. One individual from an LFS household was selected for LSUDA; no substitutes were used.

A trained interviewer visited each individual selected to participate in the LSUDA survey and administered a background questionnaire and the LSUDA test. This test consisted of a set of seven relatively easy items which were used to determine whether an individual had sufficient reading skills to undertake the more difficult items. Only those who were able to answer more than 2 of these core items continued to the main part of the test. Individuals were permitted to answer in the official language of their choice.

There were two separate administrations of the test items in the LSUDA survey. In April 1989, a pilot administration was conducted with some 1,500 individuals for the purpose of evaluating the test. Some minor adjustments to items were made, mostly to bring the French and English versions into line with each other. The final administration, to an achieved sample of approximately 9,600 individuals, was carried out in October, 1989.

4. Measuring functional literacy

Having identified certain points along the functional literacy continuum that we wanted to use as markers and having created and administered test items to serve as tests for those points, it was now possible to locate individuals along the continuum. In this section of the paper we discuss how the responses of the individuals who were tested were used to identify their functional literacy level.

4.1 Determining how to measure levels

Since we were primarily interested in the skills that underlie and generate the responses rather than the responses themselves, a measurement system that directed its attention to these underlying skills was necessary. All tests are relatively indirect measures of the skills they are directed towards; all are estimates of some underlying or latent trait.

Further, we wanted to relate the individuals' scores to the items that define the various levels. If it were possible to construct a perfect test, where the items were precise measures, it would be easy to relate individuals to levels; level 1 individuals would be unable to answer any level 2 items, level 2 individuals would answer all level 2 items, but no level 3 items. No one knows how to construct such tests, particularly tests of skills that we are still learning about, such as literacy. In the real world, level 2 people are likely to miss a few level 2 items and to answer correctly a few level 3 items. Each of the levels, since each is a point along a continuum, includes a range of abilities. The kind of scoring system that

is needed is one that relates the pattern of performance on the test to the defining items.

Item response theory (IRT) (Hambleton, 1989) provides an approach to measurement that defines individual ability in terms of the difficulty of test tasks which that individual can perform. IRT calculates for each item an estimate of its difficulty on a numerical scale and an estimate of an individual's ability using the same numerical scale, commonly a scale that ranges from 0-500. The item difficulties and individual abilities are defined in terms of each other. Briefly, an item's difficulty can be defined as the level of individual ability needed to have a certain chance of answering the item correctly; similarly, an individual's ability is defined as the level of difficulty of items which that individual has a certain chance of answering correctly. Because we are interested in a rigorous and realistic standard, we have defined that chance as 80%².

4.2 Assigning individuals to levels

In this section we discuss the use of IRT in the assignment of individuals to levels in the LSUDA study.

4.2.1 Checking reliability

Before the IRT procedures can be applied, it must be determined that the test meets standard test criteria for reliability. The reliability for the document scale items on LSUDA was .912³, quite satisfactory for a 34 item test. No single item had a major influence on the reliability.

4.2.2 Grouping items by difficulty

Once IRT difficulty scores for the items had been calculated the parameters of each level could be determined. Simply put, the items were ordered according to their difficulty score, as in Appendix A. The level for which we had designed each item was noted and the items were grouped into levels following the original expectation. In pilot trials not every item turned out to be in the expected level; each item that did not group as expected was examined to determine whether our analysis of it had been wrong or whether the item needed to be revised. Only a few items needed revision and this was largely because the French and English versions ended up at different levels. Where this happened, our examination revealed significant differences in the phrasing or the presentation of the item in the two languages or in the text used in the different versions. These differences were corrected for the final instrument.

At the same time a cluster analysis program was run to group the items by statistical similarity. This type of analysis uses several statistical tests to find the most natural groupings of objects, in this case the groupings of the test items. The item groups, or clusters, derived from this analysis matched those from the theory-

driven examination. Because of this convergence of evidence for the levels, our confidence that we have identified them properly has been strengthened.

The analysis of the item difficulties yielded the grouping in Appendix A. In short, we did the following to group the items by functional literacy level:

- 1) calculated difficulty scores for each item;
- 2) grouped the items by expected level and determined whether a grouping by difficulty scores fit the intended design;
- 3) performed a statistical grouping procedure (cluster analysis) to verify the model-based procedure in 2);
- 4) Because 2) and 3) worked satisfactorily, the statistical parameters of each level were determined from the difficulty properties of items at that level.

Thus level 2 could be determined to encompass ability scores from 160 (the easiest level 2 item) to 202 (the most difficult level 2 item). Similarly, level 3 encompasses 207 to 243; and level 4, 253 and above. This, of course, leaves small uncovered areas (194-200, 233-245). These were arbitrarily divided at even numbers so that finally:

Level 1:	Below 150
Level 2:	150-204
Level 3:	205-244
Level 4:	245 and over

4.2.3 Determining an individual's level

It was now relatively simple to determine an individual's level. Since the individual's score is the difficulty of the most difficult item that the individual has an 80% chance of answering correctly, we can use the item levels as the individual levels. So any individual whose score is less than 150 is at level 1; any individual whose score is over 205 but not over 245 is at level 3, etc.

Under this approach the technical definition of each level for individuals is:

Level 1:	Individuals who have a less than 80% probability of answering an item with a difficulty of 150 or higher.
Level 2:	Individuals who have at least an 80% probability of answering an item with a difficulty of 150 and a less than 80% probability of answering an item with a difficulty of 205.

Level 3: Individuals who have at least an 80% probability of answering an item with a difficulty of 205 and a less than 80% probability of answering an item with a difficulty of 245.

Level 4: Individuals who have at least an 80% probability of answering an item with a difficulty of 245.

Because an individual's score is based on the total pattern of answers, not just on those of a particular level, it is possible, as noted above, that some level 2 individuals will answer some level 3 items, but they will not do so consistently. Thus, an individual's level is the highest level at which she/he can perform consistently.

In general, of course, the levels are closely tied to number correct. For example, 98% of the level 2 individuals answered fewer than 21 items and 97% of the level 3 individuals answered 20 or more correctly. No level 3 individual answered fewer than 19 and no level 2 more than 22. The level 3 individuals with only 20 correct missed a few level 2 items while the level 2 individuals with 21 correct (34 individuals) answered a few level 3 items, but these appear to be simply guesses that panned out. IRT scoring looks for overall consistency and discounts, but does not ignore correct answers that are out of the pattern. Simple item correct scoring is more likely than IRT to reward an individual for a lucky guess and penalize an individual for a casual mistake. The ability to have some control over guessing and accidental errors is one of the advantages of IRT scoring procedures.

5. Conclusion

The evidence from the pilot and from the main administrations of the LSUDA indicates that the levels used in the study are identifiable points along the functional literacy continuum. It also indicates that the procedures used in assigning levels to individuals worked well. The national distribution of individuals by level in the main survey is a reasonable one (see Table 2) and that, too, gives us confidence that the levels and procedures were appropriate.

Table 2 Percentage distribution of Canadian adults aged 16-69 by level of reading skills.
LEVEL

1	6.6%
2	9.4%
3	22.1%
4	61.9%

More detailed breakdowns of the survey results, by province and background information, are available in other reports on the project. These, too, confirm the

reasonableness of the levels and procedures. Just how useful the levels are, however, depends on the use of the results by literacy practitioners and on how they are used in analyses of the data by other researchers. We feel that they have enabled us to look closely at literacy abilities in Canada and we are confident that they have yielded useful and dependable information.

6. Notes

1. Because functional literacy is a continuum, individuals within a level have a range of skills. Using a test with items with a range of difficulty would also allow analysis within a level.
2. This is the standard used in the survey of Young Adult Literacy in the United States (Kirsch and Jungeblut, 1986).
3. Reliability is a standard measure of test quality. The measure used here, Cronbach's alpha, is a measure of whether all the items on the test are measuring the same skill. Reliabilities range from 0.0, which indicates no consistency of measurement, to 1.00 which indicates complete consistency. Tests with reliabilities over .9 are regarded as highly consistent and acceptable measuring instruments.

7. Bibliography

Guthrie, J. T. (1988). Locating information in documents: examination of a cognitive model. *Reading Research Quarterly*, 23, 178-199.

Guthrie, J. T., & Kirsch, I. S. (1987). Distinctions between reading comprehension and locating information in text. *Journal of Educational Psychology*, 79, 220-227.

Hambleton, R. K. (1989). Principles and selected applications of Item Response Theory. In R.L. Linn (Ed.) *Educational Measurement, 3rd Edition* (pp. 147-200). NY American Council of Education.

Jones, S. (1989). *Design of test items for the survey of literacy skills used in daily activities*. Ottawa: Statistics Canada.

Jones, S., Satin, A., Kelly, K., & Montigny, G. (1990). *National literacy skill assessment: the Canadian experience*. (Unpublished paper). Ottawa: Statistics Canada.

Kelly, K., Murray, S., & Satin, A. (1988). *A national literacy skills assessment* (Planning Report, Special Surveys Group). Ottawa: Statistics Canada.

Kirsch, I. S., & Guthrie, J. T. (1981). The concept and measurement of functional literacy. *Reading Research Quarterly*, 13, 485-507.

Kirsch, I. S., & Jungeblut, A. (1986). *Literacy: profiles of America's young adults* (Final Report, 16-PL-01). Princeton, NJ: The National Assessment of Educational Progress.

Mikulecky, L. (1985, March). *Literacy task analysis: defining and measuring occupational literacy demands*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Jones, S., Satin, A., Kelly, K., & Montigny, G. (1990). *National literacy skill assessment: the Canadian experience*. (Unpublished paper). Ottawa: Statistics Canada.

APPENDIX A
Item Difficulties and Levels

LEVEL	DIFFICULTY	DESCRIPTION
1	93	Social Insurance Card
2	160	Grocery Ad
	178	Marathon Swimmer Goal
	184	Enumeration Form
	184	School Letter Place X
	185	Building Signs
	190	Marathon Swimmer Eats What
	194	Drivers Licence
	197	Telephone Bill
	200	Financial Graph
	202	Goods to Market
3	207	Pool Schedule Family Swim
	210	Peddlar
	212	School Letter Return When
	212	Vacation Cost
	228	Eligibility Chart Health Plan
	230	Deposit Slip Cash
	233	Sandpaper Selection Extra Fine
	238	Yellow Pages Cabat
	239	Grocery Label Compare
	240	Sandpaper Selection Metal
	243	Medicine Label
	253	Deposit Slip List Cheques
4	254	Yellow Pages Railroads
	266	Order Form Address
	269	Amherst Map
	273	Order Form Quantity
	274	Eligibility Chart LTD
	276	Deposit Slip Date
	284	Pool Schedule Seniors Swim
	293	Classified Ad
	295	School Hours
	330	Aging Population Evidence
	332	Line Graph
	361	Aging Population Importance

